GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 11:11:10 ; Search time 16.02 Seconds
(without alignments)
28.222 Million cell updates/sec

Title: US-09-288-719-2
Perfect score: 75
Sequence: 1 GGGGSGGRASGGGS 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	65	86.7	643	1	K2C1_HUMAN	P04264 homo sapien
2	62	82.7	206	1	TWST_MOUSE	P26687 mus musculu
3	61	81.3	3038	1	TRIO_HUMAN	O75962 homo sapien
4	60	80.0	401	1	CSP_PLACG	P08674 plasmodium
5	60	80.0	465	1	FXD1_HUMAN	Q16676 homo sapien
6	60	80.0	465	1	FXD3_MOUSE	Q61060 mus musculu
7	60	80.0	779	1	M130_STRPU	P08472 strongyloce
8	59	78.7	558	1	ROL_HUMAN	P14866 homo sapien
9	59	78.7	593	1	K1CJ_HUMAN	P13645 homo sapien
10	58	77.3	214	1	GRP2_NICSY	P27484 nicotiana s
11	58	77.3	333	1	SIX3_MOUSE	Q62233 mus musculu
12	58	77.3	427	1	DSXF_DROME	P23022 drosophila
13	58	77.3	481	1	LORI_MOUSE	P18165 mus musculu
14	58	77.3	549	1	DSXM_DROME	P23023 drosophila
15	57	76.0	118	1	RLA3_ORYSA	P56724 oryza sativ
16	57	76.0	265	1	YPG1_ZYMMO	P30794 zymomonas m
17	57	76.0	266	1	CANS_RABIT	P06813 oryctolagus
18	57	76.0	268	1	EP34_HCMVA	P16768 human cytom
19	57	76.0	431	1	HXB3_HUMAN	P14651 homo sapien
20	57	76.0	433	1	HXB3_MOUSE	P09026 mus musculu
21	57	76.0	575	1	RM62_DROME	P19109 drosophila
22	57	76.0	684	1	EP84_HCMVA	P17151 human cytom
23	57	76.0	1454	1	KDGE_DROME	Q09103 drosophila
24	56	74.7	157	1	GRPA_MAIZE	P10979 zea mays (m
25	56	74.7	157	1	GRP_DAUCA	Q03878 daucus caro
26	56	74.7	165	1	GRP1_ORYSA	P25074 oryza sativ
27	56	74.7	168	1	GRP2_SORBI	Q99070 sorghum bic
28	56	74.7	183	1	GRP2_ORYSA	P29834 oryza sativ
29	56	74.7	202	1	TWST_HUMAN	Q15672 homo sapien
30	56	74.7	280	1	CHIA_MAIZE	P29022 zea mays (m
31	56	74.7	321	1	PUR_MOUSE	P42669 mus musculu
32	56	74.7	322	1	PUR_HUMAN	Q00577 homo sapien
33	56	74.7	385	1	RO32_XENLA	P51992 xenopus lae
34	56	74.7	401	1	HB9_HUMAN	P50219 homo sapien
35	56	74.7	440	1	FXGA_CHICK	Q98937 gallus gall
36	56	74.7	495	1	BRN1_MOUSE	P31361 mus musculu
37	56	74.7	497	1	BRN1_RAT	Q63262 rattus norv
38	56	74.7	500	1	BRN1_HUMAN	P20264 homo sapien
39	56	74.7	539	1	DOP2_DROME	Q24563 drosophila
40	56	74.7	622	1	SR68_CANFA	Q00004 canis famil
41	56	74.7	663	1	DUS8_MOUSE	O09112 mus musculu
42	56	74.7	688	1	EOMD_MOUSE	O54839 mus musculu
43	56	74.7	979	1	RFX1_HUMAN	P22670 homo sapien
44	56	74.7	1627	1	TP2B_CHICK	O42131 gallus gall
45	56	74.7	2124	1	Y192_HUMAN	Q93074 homo sapien

ALIGNMENTS

RESULT 1
K2C1_HUMAN
ID K2C1_HUMAN STANDARD; PRT; 643 AA.
AC P04264;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE KERATIN, TYPE II CYTOSKELETAL 1 (CYTOKERATIN 1) (K1) (CK 1) (67 KDA
DE CYTOKERATIN) (HAIR ALPHA PROTEIN).
GN KRT1 OR KRTA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85166239; PubMed=2580302;
RA Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;
RT "Structure of a gene for the human epidermal 67-kDa keratin.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA Whittock N.V., Eady R.A., McGrath J.A.;
RT "Genomic organization of the human keratin 1 gene.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP PRELIMINARY SEQUENCE OF 151-643 FROM N.A.
RX MEDLINE=85207740; PubMed=2581964;
RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,
RA Roop D.R.;
RT "Amino acid sequences of mouse and human epidermal type II keratins
RT of Mr 67,000 provide a systematic basis for the structural and
RT functional diversity of the end domains of keratin intermediate
RT filament subunits.";
RL J. Biol. Chem. 260:7142-7149(1985).
RN [4]
RP REVISIONS, AND VARIANT EHK PRO-160.
RX MEDLINE=92386601; PubMed=1381288;
RA Chipev C.C., Korge B.P., Markova N., Bale S.J., Digiovanna J.J.,
RA Compton J.G., Steinert P.M.;
RT "A leucine-->proline mutation in the H1 subdomain of keratin 1
RT causes epidermolytic hyperkeratosis.";
RL Cell 70:821-828(1992).
RN [5]
RP VARIANTS EHK GLY-154; SER-187 AND PRO-192.
RX MEDLINE=94117869; PubMed=7507151;
RA Yang J.-M., Chipev C.C., Digiovanna J.J., Bale S.J., Marekov L.N.,
RA Steinert P.M., Compton J.G.;
RT "Mutations in the H1 and 1A domains in the keratin 1 gene in
RT epidermolytic hyperkeratosis.";
RL J. Invest. Dermatol. 102:17-23(1994).
RN [6]
RP VARIANTS EHK PRO-185 AND SER-187.
RX MEDLINE=94117870; PubMed=7507152;
RA McLean W.H.I., Eady R.A., Dopping-Hepenstal P.J., McMillan J.R.,
RA Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,
RA Morley S.M.;

```
RT   "Mutations in the rod 1A domain of keratins 1 and 10 in bullous
RT   congenital ichthyosiform erythroderma (BCIE).";
RL   J. Invest. Dermatol. 102:24-30(1994).
RN   [7]
RP   VARIANT EHK GLN-489.
RX   MEDLINE=92376531; PubMed=1380725;
RA   Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,
RA   Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;
RT   "Mutations in the rod domains of keratins 1 and 10 in epidermolytic
RT   hyperkeratosis.";
RL   Science 257:1128-1130(1992).
RN   [8]
RP   VARIANT ALLELE 1B.
RX   MEDLINE=93107743; PubMed=1281859;
RA   Korge B.P., Compton J.G., Steinert P.M., Mischke D.;
RT   "The two size alleles of human keratin 1 are due to a deletion in the
RT   glycine-rich carboxyl-terminal V2 subdomain.";
RL   J. Invest. Dermatol. 99:697-702(1992).
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC       KERATIN 1 IS GENERALLY ASSOCIATED WITH KERATIN 10.
CC   -!- TISSUE SPECIFICITY: THE SOURCE OF THIS PROTEIN IS NEONATAL
CC       FORESKIN. THE 67-KDA TYPE II KERATINS ARE EXPRESSED IN TERMINALLY
CC       DIFFERENTIATING EPIDERMIS.
CC   -!- POLYMORPHISM: THERE ARE TWO SIZE VARIANTS OF KRT1, TERMED 1A AND
CC       1B WITH ALLELIC FREQUENCIES OF 0.61 AND 0.39. 1B LACKS 7 LACKS
CC       7 RESIDUES COMPARED TO 1A.
CC   -!- DISEASE: DEFECTS IN KRT10 AND KRT1 ARE THE CAUSE OF EPIDERMOLYTIC
CC       HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL
CC       ICHTHYOSIFORM ERYTHRODERMA (BIE)); A HEREDITARY SKIN DISORDER
CC       CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
CC       CORNEUM.
CC   -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC       MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II
CC       (NEUTRAL TO BASIC; 56-70 KDA) [K1 TO K8].
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF237621; AAF60327.1; -.
DR   EMBL; M98776; AAB47721.1; -.
DR   PIR; A02950; KRHU2.
DR   PIR; A22940; A22940.
DR   HSSP; P05412; 1FOS.
DR   AARHUS/GHENT-2DPAGE; 4606; NEPHGE.
DR   MIM; 139350; -.
DR   MIM; 113800; -.
DR   INTERPRO; IPR001664; -.
DR   INTERPRO; IPR003054; -.
DR   PFAM; PF00038; filament; 1.
DR   PRINTS; PR01276; TYPE2KERATIN.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin;
KW   Disease mutation; Polymorphism; Phosphorylation.
FT   INIT_MET      0      0
FT   DOMAIN        1    178       HEAD.
FT   DOMAIN      179    488       ROD.
FT   DOMAIN      489    643       TAIL.
FT   DOMAIN      179    214       COIL 1A.
FT   DOMAIN      215    233       LINKER 1.
FT   DOMAIN      234    325       COIL 1B.
FT   DOMAIN      326    349       LINKER 12.
FT   DOMAIN      350    488       COIL 2.
FT   SITE        432    432       STUTTER.
FT   DOMAIN        1    150       GLY/PHE/SER-RICH.
FT   DOMAIN      501    640       GLY/SER-RICH.
FT   MOD_RES      65     65       PHOSPHORYLATION (BY SIMILARITY).
FT   VARIANT     154    154       V -> G (IN EHK).
FT                                /FTId=VAR_003853.
FT   VARIANT     160    160       L -> P (IN EHK).
FT                                /FTId=VAR_003854.
FT   VARIANT     185    185       S -> P (IN EHK).
FT                                /FTId=VAR_003855.
FT   VARIANT     187    187       N -> S (IN EHK).
FT                                /FTId=VAR_003856.
FT   VARIANT     192    192       S -> P (IN EHK).
FT                                /FTId=VAR_003857.
FT   VARIANT     311    311       I -> V.
FT                                /FTId=VAR_003858.
FT   VARIANT     329    329       I -> T.
FT                                /FTId=VAR_003859.
FT   VARIANT     357    357       N -> Y.
FT                                /FTId=VAR_003860.
FT   VARIANT     489    489       E -> Q (IN EHK).
FT                                /FTId=VAR_003861.
FT   VARIANT     536    536       G -> C.
FT                                /FTId=VAR_003862.
FT   VARIANT     632    632       R -> K.
FT                                /FTId=VAR_003863.
FT   VARIANT     559    565       MISSING (IN ALLELE 1B).
FT                                /FTId=VAR_003864.
SQ   SEQUENCE   643 AA;  65886 MW;  DF945DC462257850 CRC64;


  Query Match             86.7%;  Score 65;  DB 1;  Length 643;
  Best Local Similarity   85.7%;  Pred. No. 1.3;
  Matches   12;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           ||| |||| |||||
Db     594 GGGSSGGRGSGGGS 607


RESULT   2
TWST_MOUSE
ID   TWST_MOUSE     STANDARD;      PRT;   206 AA.
AC   P26687;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   TWIST RELATED PROTEIN (M-TWIST).
GN   TWIST.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91122450; PubMed=1840517;
RA   Wolf C., Thisse C., Stoetzel C., Thisse B., Gerlinger P.,
RA   Perrin-Schmitt F.;
RT   "The M-twist gene of Mus is expressed in subsets of mesodermal cells
RT   and is closely related to the Xenopus X-twi and the Drosophila twist
RT   genes.";
RL   Dev. Biol. 143:363-373(1991).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=98001585; PubMed=9343420;
RA   Hamamori Y., Wu H.Y., Sartorelli V., Kedes L.;
RT   "The basic domain of myogenic basic helix-loop-helix (bHLH) proteins
RT   is the novel target for direct inhibition by another bHLH protein,
RT   Twist.";
RL   Mol. Cell. Biol. 17:6563-6573(1997).
CC   -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR, WHICH SEEMS TO BE
CC       INVOLVED IN THE NEGATIVE REGULATION OF CELLULAR DETERMINATION AND
CC       IN THE DIFFERENTIATION OF SEVERAL LINEAGES INCLUDING MYOGENESIS,
CC       OSTEOGENESIS, AND NEUROGENESIS. INHIBITS MYOGENESIS BY
CC       SEQUESTRATING E PROTEINS, INHIBITING TRANS-ACTIVATION BY MEF2, AND
CC       INHIBITING DNA-BINDING BY MYOD THROUGH PHYSICAL INTERACTION. THIS
CC       INTERACTION PROBABLY INVOLVES THE BASIC DOMAINS OF BOTH PROTEINS.
CC   -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC       BHLH PROTEIN. HOMODIMER.
```

```
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: SUBSET OF MESODERMAL CELLS.
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC       TRANSCRIPTION FACTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63649; AAA40514.1; -.
DR   EMBL; M63650; AAA40515.1; -.
DR   HSSP; P10085; 1MDY.
DR   TRANSFAC; T01635; -.
DR   MGD; MGI:98872; TWIST.
DR   INTERPRO; IPR001092; -.
DR   INTERPRO; IPR003015; -.
DR   PFAM; PF00010; HLH; 1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW   Differentiation; Developmental protein; Nuclear protein; DNA-binding;
KW   Transcription regulation.
FT   DOMAIN       80    102       GLY-RICH.
FT   DNA_BIND    112    124       BASIC DOMAIN.
FT   DOMAIN      125    164       HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT   VARIANT      36     36       A -> R (IN CDNA).
FT   VARIANT      91     91       G -> P (IN CDNA).
SQ   SEQUENCE   206 AA;  21198 MW;  618AD8E9BE87C555 CRC64;


  Query Match             82.7%;  Score 62;  DB 1;  Length 206;
  Best Local Similarity   78.6%;  Pred. No. 1;
  Matches   11;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           |||| || :|||||
Db      90 GGGGGGGSSSGGGS 103


RESULT   3
TRIO_HUMAN
ID   TRIO_HUMAN     STANDARD;      PRT;  3038 AA.
AC   O75962; Q13458;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TRIPLE FUNCTIONAL DOMAIN PROTEIN (PTPRF INTERACTING PROTEIN).
GN   TRIO.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=FIBROBLAST;
RX   MEDLINE=96224308; PubMed=8643598;
RA   Debant A., Serra-Pages C., Seipel K., O'Brien S., Tang M., Park S.-H.,
RA   Streuli M.;
RT   "The multidomain protein Trio binds the LAR transmembrane tyrosine
RT   phosphatase, contains a protein kinase domain, and has separate rac-
RT   specific and rho-specific guanine nucleotide exchange factor
RT   domains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:5466-5471(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Streuli M.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.
RX   MEDLINE=99005194; PubMed=9790533;
RA   Liu X., Wang H., Eberstadt M., Schnuchel A., Olejniczak E.T.,
RA   Meadows R.P., Schkeryantz J.M., Janowick D.A., Harlan J.E.,
RA   Harris E.A.S., Staunton D.E., Fesik S.W.;
RT   "NMR structure and mutagenesis of the N-terminal Dbl homology domain
RT   of the nucleotide exchange factor Trio.";
RL   Cell 95:269-277(1998).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=99276567; PubMed=10341202;
RA   Seipel K., Medley Q.G., Kedersha N.L., Zhang X.A., O'Brien S.P.,
RA   Serra-Pages C., Hemler M.E., Streuli M.;
RT   "Trio amino-terminal guanine nucleotide exchange factor domain
RT   expression promotes actin cytoskeleton reorganization, cell migration
RT   and anchorage-independent cell growth.";
RL   J. Cell Sci. 112:1825-1834(1999).
CC   -!- FUNCTION: PROMOTES THE EXCHANGE OF GDP BY GTP. TOGETHER WITH
CC       LEUCOCYTE ANTIGEN-RELATED (LAR) PROTEIN, IT COULD PLAY A ROLE IN
CC       COORDINATING CELL-MATRIX AND CYTOSKELETAL REARRANGEMENTS NECESSARY
CC       FOR CELL MIGRATION AND CELL GROWTH.
CC   -!- SUBUNIT: INTERACT TO FORM A COMPLEX WITH LEUKOCYTE ANTIGEN RELATED
CC       PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, SKELETAL MUSCLE,
CC       BRAIN, PANCREAS, PLACENTA, LIVER, KIDNEY AND LUNG.
CC   -!- DOMAIN: THE N-TERMINAL DBL/GEF DOMAIN SPECIFICALLY CATALYZES
CC       NUCLEOTIDE EXCHANGE FOR RAC1, LEADING TO THE ACTIVATION OF JUN
CC       KINASE AND THE PRODUCTION OF MEMBRANE RUFFLES. THE SECOND DBL/GEF
CC       DOMAIN IS AN EXCHANGE FACTOR FOR RHOA AND INDUCES THE FORMATION OF
CC       STRESS FIBERS.
CC   -!- PTM: SERINE PHOSPHORYLATED.
CC   -!- SIMILARITY: BELONGS TO THE RHO/RAC GEF FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 DBL-HOMOLOGY DOMAINS (DH).
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 SPECTRIN-LIKE DOMAINS.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SER/THR
CC       FAMILY OF PROTEIN KINASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF091395; AAC43042.1; -.
DR   EMBL; U42390; AAC34245.1; -.
DR   HSSP; P04002; 1ATF.
DR   MIM; 601893; -.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00169; PH; 2.
DR   PFAM; PF00069; pkinase; 1.
DR   PFAM; PF00621; RhoGEF; 2.
DR   PFAM; PF00018; SH3; 1.
DR   PFAM; PF00435; spectrin; 7.
DR   PROSITE; PS50003; PH_DOMAIN; 2.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Guanine-nucleotide releasing factor; Phosphorylation; Repeat;
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Immunoglobulin domain; SH3 domain; Alternative splicing.
FT   DOMAIN       252   1157       4 X SPECTRIN-LIKE REPEATS.
FT   REPEAT       252    359       1.
FT   REPEAT       479    585       2.
FT   REPEAT       819    925       3.
FT   REPEAT      1050   1157       4.
FT   DOMAIN      1240   1393       DH 1.
FT   DOMAIN      1421   1532       PH 1.
FT   DOMAIN      1597   1653       SH3.
FT   DOMAIN      1917   2091       DH 2.
```

```
FT   DOMAIN     2098   2212       PH 2.
FT   DOMAIN     2630   2707       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN     2737   2993       PROTEIN KINASE.
FT   DOMAIN      656    659       POLY-GLN.
FT   DOMAIN     1786   1791       POLY-SER.
FT   DOMAIN     2233   2248       POLY-GLY.
FT   DOMAIN     2241   2253       POLY-GLY.
FT   DOMAIN     2486   2492       POLY-SER.
FT   DISULFID   2637   2700       POTENTIAL.
FT   NP_BIND    2743   2751       ATP (BY SIMILARITY).
FT   BINDING    2766   2766       ATP (BY SIMILARITY).
FT   ACT_SITE   2856   2856       BY SIMILARITY.
FT   MUTAGEN    1240   1240       E->A: 50% DECREASE IN NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   MUTAGEN    1244   1244       T->A: 40% DECREASE IN NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   MUTAGEN    1330   1330       N->A: NO CHANGE IN NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   MUTAGEN    1367   1367       V->A: 90% DECREASE IN NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   MUTAGEN    1368   1368       Q->A: 80% DECREASE IN NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   MUTAGEN    1369   1369       R->A: 80% DECREASE IN NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   MUTAGEN    1371   1371       T->A: 80% DECREASE IN NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   MUTAGEN    1372   1372       K->A: LOSS OF NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   MUTAGEN    1375   1375       L->A: 40% DECREASE IN NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   MUTAGEN    1378   1378       K->A: NO CHANGE IN NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   MUTAGEN    1379   1379       E->A: 30% DECREASE IN NUCLEOTIDE EXCHANGE
FT                                ACTIVITY.
FT   VARSPLIC   2242   2308       GSGGGGGSGGGGAPSGGSGHSGGPSSCGGAPSTSRSRPSRI
FT                                PQPVRHHPPVLVSSAASSQAEADKMS -> AAAGVGAAAAA
FT                                GPPVAAAATVAAPAAAAAPPARAGAGPPGSPSLSDTTPPCW
FT                                SPLQPRARQRQTRCQ (IN ISOFORM 2).
FT   VARSPLIC   2309   3038       MISSING (IN ISOFORM 2).
SQ   SEQUENCE   3038 AA;  341611 MW;  28620F3B513EB74B CRC64;


  Query Match             81.3%;  Score 61;  DB 1;  Length 3038;
  Best Local Similarity   78.6%;  Pred. No. 13;
  Matches   11;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           |||||||   ||||
Db    2236 GGGGSGGSGGGGGS 2249


RESULT   4
CSP_PLACG
ID   CSP_PLACG      STANDARD;      PRT;   401 AA.
AC   P08674;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).
OS   Plasmodium cynomolgi (strain Gombak).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87102878; PubMed=3802196;
RA   Galinski M.R., Arnot D.E., Cochrane A.H., Barnwell J.W.,
RA   Nussenzweig R.S., Enea V.;
RT   "The circumsporozoite gene of the Plasmodium cynomolgi complex.";
RL   Cell 48:311-319(1987).
CC   -!- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC       SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC       MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC       VERTEBRATE HOST).
CC   -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC       ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC       WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC   -!- SIMILARITY: CONTAINS 1 TYPE-1 TSP REPEAT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15100; AAA29536.1; -.
DR   PIR; E26255; OZZQAC.
DR   INTERPRO; IPR000884; -.
DR   INTERPRO; IPR003067; -.
DR   PFAM; PF00090; tsp_1; 1.
DR   PRINTS; PR01303; CRCMSPRZOITE.
KW   Malaria; Sporozoite; Repeat; Signal.
FT   SIGNAL        1     19       PROBABLE.
FT   CHAIN        20    401       CIRCUMSPOROZOITE PROTEIN.
FT   DOMAIN       98    278       17 X 11 AA TANDEM REPEATS OF [DG]-G-A-A-
FT                                A-A-G-G-G-G-N.
SQ   SEQUENCE   401 AA;  36664 MW;  57D666268238503E CRC64;


  Query Match             80.0%;  Score 60;  DB 1;  Length 401;
  Best Local Similarity   71.4%;  Pred. No. 2.9;
  Matches   10;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           ||||:|| |:|||:
Db     256 GGGGNGGAAAGGGN 269


RESULT   5
FXD1_HUMAN
ID   FXD1_HUMAN     STANDARD;      PRT;   465 AA.
AC   Q16676; Q12949;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   FORKHEAD BOX PROTEIN D1 (FORKHEAD-RELATED PROTEIN FKHL8) (FORKHEAD-
DE   RELATED TRANSCRIPTION FACTOR 4) (FREAC-4).
GN   FOXD1 OR FKHL8 OR FREAC4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96355467; PubMed=8702877;
RA   Ernstsson S., Pierrou S., Hulander M., Cederberg A., Hellqvist M.,
RA   Carlsson P., Enerback S.;
RT   "Characterization of the human forkhead gene FREAC-4. Evidence for
RT   regulation by Wilms' tumor suppressor gene (WT-1) and p53.";
RL   J. Biol. Chem. 271:21094-21099(1996).
RN   [2]
RP   SEQUENCE OF 120-225 FROM N.A.
RX   MEDLINE=95045392; PubMed=7957066;
RA   Pierrou S., Hellqvist M., Samuelsson L., Enerbaeck S., Carlsson P.;
RT   "Cloning and characterization of seven human forkhead proteins:
RT   binding site specificity and DNA bending.";
RL   EMBO J. 13:5002-5012(1994).
CC   -!- FUNCTION: BINDING OF FREAC-3 AND FREAC-4 TO THEIR COGNATE SITES
CC       RESULTS IN BENDING OF THE DNA AT AN ANGLE OF 80-90 DEGREES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U59832; AAC50661.1; -.
DR   EMBL; U59831; AAC50660.1; -.
DR   EMBL; U13222; AAA92039.1; -.
DR   MIM; 601091; -.
DR   INTERPRO; IPR001766; -.
DR   PFAM; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   PROSITE; PS00657; FORK_HEAD_1; 1.
DR   PROSITE; PS00658; FORK_HEAD_2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein.
FT   DOMAIN       26     34       POLY-GLU.
FT   DOMAIN       39     43       POLY-GLY.
FT   DOMAIN       52     57       POLY-ARG.
FT   DOMAIN       69     72       POLY-GLU.
FT   DOMAIN       73     76       POLY-ASP.
FT   DOMAIN       97    113       POLY-GLY.
FT   DNA_BIND    124    215       FORK-HEAD.
FT   DOMAIN      231    234       POLY-ALA.
FT   DOMAIN      252    256       POLY-ALA.
FT   DOMAIN      259    266       POLY-PRO.
FT   DOMAIN      293    303       POLY-ALA.
FT   DOMAIN      309    315       POLY-PRO.
FT   DOMAIN      375    390       POLY-ALA.
FT   DOMAIN      428    434       POLY-ALA.
SQ   SEQUENCE   465 AA;  46140 MW;  D3E7854909CCBFAE CRC64;


  Query Match             80.0%;  Score 60;  DB 1;  Length 465;
  Best Local Similarity   78.6%;  Pred. No. 3.3;
  Matches   11;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 GGGGSGGRASGGGS 14
          |||| || | ||||
Db    101 GGGGGGGGAGGGGS 114


RESULT   6
FXD3_MOUSE
ID   FXD3_MOUSE     STANDARD;      PRT;   465 AA.
AC   Q61060;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   FORKHEAD BOX PROTEIN D3 (HNF3/FH TRANSCRIPTION FACTOR GENESIS)
DE   (HEPATOCYTE NUCLEAR FACTOR 3 FORKHEAD HOMOLOG 2) (HFH-2).
GN   FOXD3 OR HFH2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hromas R.A., Costa R.H., Xu D., Sutton J.L.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Labosky P.A., Kaestner K.H.;
RT   "The winged helix trancription factor Hfh2 is expressed in neural
RT   crest and spinal cord during mouse development.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U41047; AAA87569.1; -.
DR   EMBL; AF067421; AAC28352.1; -.
DR   MGD; MGI:1347473; FOXD3.
DR   INTERPRO; IPR001766; -.
DR   PFAM; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   PROSITE; PS00657; FORK_HEAD_1; 1.
DR   PROSITE; PS00658; FORK_HEAD_2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
KW   DNA-binding; Nuclear protein; Transcription regulation.
FT   DOMAIN       106    113       POLY-GLY.
FT   DNA_BIND     131    225       FORK-HEAD.
FT   DOMAIN       252    257       POLY-ALA.
FT   DOMAIN       265    270       POLY-ALA.
FT   DOMAIN       275    281       POLY-ALA.
FT   DOMAIN       380    393       POLY-GLY.
FT   DOMAIN       385    399       POLY-GLY.
FT   DOMAIN       447    457       POLY-ALA.
SQ   SEQUENCE   465 AA;  47092 MW;  6F8B5BB3D8C7564D CRC64;


  Query Match             80.0%;  Score 60;  DB 1;  Length 465;
  Best Local Similarity   84.6%;  Pred. No. 3.3;
  Matches   11;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 GGGGSGGRASGGG 13
          ||||||| | |||
Db    380 GGGGSGGGAGGGG 392


RESULT   7
M130_STRPU
ID   M130_STRPU     STANDARD;      PRT;   779 AA.
AC   P08472;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-MAY-1991 (Rel. 18, Last annotation update)
DE   MESENCHYME-SPECIFIC CELL SURFACE GLYCOPROTEIN PRECURSOR (MSP130).
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90110195; PubMed=2295637;
RA   Parr B.A., Parks A.L., Raff R.A.;
RT   "Promoter structure and protein sequence of msp130, a lipid-anchored
RT   sea urchin glycoprotein.";
RL   J. Biol. Chem. 265:1408-1413(1990).
RN   [2]
RP   SEQUENCE OF 542-779 FROM N.A.
RX   MEDLINE=87191419; PubMed=3569664;
RA   Leaf D.S., Anstrom J.A., Chin J.E., Harkey M.A., Showman R.M.,
RA   Raff R.A.;
RT   "Antibodies to a fusion protein identify a cDNA clone encoding
RT   msp130, a primary mesenchyme-specific cell surface protein of the sea
RT   urchin embryo.";
RL   Dev. Biol. 121:29-40(1987).
CC   -!- FUNCTION: NOT KNOWN. COULD BE INVOLVED IN MESENCHYME CELL
CC       MIGRATION, ADHESION, FUSION, OR SPICULE FORMATION.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (PROBABLE).
CC   -!- TISSUE SPECIFICITY: RESTRICTED TO THE PRIMARY MESENCHYME CELL
CC       LINEAGE.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31750; AAA30065.1; -.
DR   EMBL; M31751; AAA30066.1; -.
DR   EMBL; M16457; AAA30064.1; -.
DR   PIR; A35006; A35006.
KW   Glycoprotein; Signal; Repeat; GPI-anchor.
FT   SIGNAL        1     15       POTENTIAL.
FT   CHAIN        16    779       MESENCHYME-SPECIFIC CELL SURFACE
FT                                GLYCOPROTEIN.
FT   DOMAIN       39     70       GLY-RICH.
FT   DOMAIN      262    350       GLY-RICH.
FT   CARBOHYD    203    203       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    451    451       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    609    609       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   779 AA;  81006 MW;  7D522D8348928316 CRC64;


  Query Match             80.0%;  Score 60;  DB 1;  Length 779;
  Best Local Similarity   76.9%;  Pred. No. 5.2;
  Matches   10;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGG 13
           ||||:|||  |||
Db      57 GGGGAGGRGGGGG 69


RESULT   8
ROL_HUMAN
ID   ROL_HUMAN      STANDARD;      PRT;   558 AA.
AC   P14866;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN L (HNRNP L).
GN   HNRPL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90078296; PubMed=2687284;
RA   Pinol-Roma S., Swanson M.S., Gall J.G., Dreyfuss G.;
RT   "A novel heterogeneous nuclear RNP protein with a unique distribution
RT   on nascent transcripts.";
RL   J. Cell Biol. 109:2575-2587(1989).
RN   [2]
RP   PARTIAL SEQUENCE.
RC   TISSUE=KERATINOCYTES;
RX   MEDLINE=93162043; PubMed=1286667;
RA   Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA   Vandekerckhove J.;
RT   "Microsequences of 145 proteins recorded in the two-dimensional gel
RT   protein database of normal human epidermal keratinocytes.";
RL   Electrophoresis 13:960-969(1992).
CC   -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE HETEROGENOUS NUCLEAR
CC       RIBONUCLEOPROTEIN (HNRNP) COMPLEXES WHICH PROVIDE THE SUBSTRATE
CC       FOR THE PROCESSING EVENTS THAT PRE-MRNAS UNDERGO BEFORE BECOMING
CC       FUNCTIONAL, TRANSLATABLE MRNAS IN THE CYTOPLASM. L IS ASSOCIATED
CC       WITH MOST NASCENT TRANSCRIPTS INCLUDING THOSE OF THE LANDMARK
CC       GIANT LOOPS OF AMPHIBIAN LAMPBRUSH CHROMOSOMES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOPLASM.
CC   -!- PTM: SEVERAL ISOELECTRIC FORMS OF THE L PROTEIN ARE PROBABLY THE
CC       RESULTS OF POSTTRANSLATIONAL MODIFICATIONS.
CC   -!- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RRM).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16135; CAA34261.1; -.
DR   PIR; A33616; A33616.
DR   HSSP; P19339; 2SXL.
DR   SWISS-2DPAGE; P14866; HUMAN.
DR   AARHUS/GHENT-2DPAGE; 1505; IEF.
DR   AARHUS/GHENT-2DPAGE; 4602; NEPHGE.
DR   MIM; 603083; -.
DR   MIM; 164021; -.
DR   INTERPRO; IPR000504; -.
DR   PFAM; PF00076; rrm; 3.
DR   PROSITE; PS50102; RRM; 3.
KW   Nuclear protein; RNA-binding; Ribonucleoprotein; Repeat.
FT   DOMAIN        8     58       GLY-RICH.
FT   DOMAIN       72    552       REPEAT-RICH REGION.
FT   REPEAT       72    143
FT   REPEAT      163    238
FT   REPEAT      352    424
FT   REPEAT      472    552
FT   DOMAIN      335    342       POLY-PRO.
FT   SIMILAR      63    143       TO RNA-BINDING REGION OF HNRNP AND SNRNP
FT                                PROTEINS.
FT   SIMILAR     155    237       TO RNA-BINDING REGION OF HNRNP AND SNRNP
FT                                PROTEINS.
SQ   SEQUENCE   558 AA;  60187 MW;  395E5A04B14C848D CRC64;


  Query Match             78.7%;  Score 59;  DB 1;  Length 558;
  Best Local Similarity   78.6%;  Pred. No. 5;
  Matches   11;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           |||| |||  ||||
Db       8 GGGGGGGRYYGGGS 21


RESULT   9
K1CJ_HUMAN
ID   K1CJ_HUMAN     STANDARD;      PRT;   593 AA.
AC   P13645;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   KERATIN, TYPE I CYTOSKELETAL 10 (CYTOKERATIN 10) (K10) (CK 10).
GN   KRT10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89125611; PubMed=2464696;
RA   Rieger M., Franke W.W.;
RT   "Identification of an orthologous mammalian cytokeratin gene. High
RT   degree of intron sequence conservation during evolution of human
RT   cytokeratin 10.";
RL   J. Mol. Biol. 204:841-856(1988).
RN   [2]
RP   SEQUENCE OF 130-593 FROM N.A.
RX   MEDLINE=88122104; PubMed=2448602;
RA   Darmon M.Y., Semat A., Darmon M.C., Vasseur M.;
RT   "Sequence of a cDNA encoding human keratin No 10 selected according
RT   to structural homologies of keratins and their tissue-specific
RT   expression.";
RL   Mol. Biol. Rep. 12:277-283(1987).
RN   [3]
RP   SEQUENCE OF 197-593 FROM N.A.
RX   MEDLINE=92339897; PubMed=1378806;
RA   Tkachenko A.V., Buchman V.L., Bliskovsky V.V., Shvets Y.P.,
RA   Kisselev L.L.;
RT   "Exons I and VII of the gene (Ker10) encoding human keratin 10
RT   undergo structural rearrangements within repeats.";
```

```
RL   Gene 116:245-251(1992).
RN   [4]
RP   SEQUENCE OF 180-184 AND 577-589.
RC   TISSUE=KERATINOCYTES;
RX   MEDLINE=93162043; PubMed=1286667;
RA   Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA   Vandekerckhove J.;
RT   "Microsequences of 145 proteins recorded in the two-dimensional gel
RT   protein database of normal human epidermal keratinocytes.";
RL   Electrophoresis 13:960-969(1992).
RN   [5]
RP   VARIANTS EHK HIS-156 AND SER-161.
RX   MEDLINE=92376531; PubMed=1380725;
RA   Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,
RA   Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;
RT   "Mutations in the rod domains of keratins 1 and 10 in epidermolytic
RT   hyperkeratosis.";
RL   Science 257:1128-1130(1992).
RN   [6]
RP   VARIANT EHK HIS-156.
RX   MEDLINE=92386600; PubMed=1381287;
RA   Cheng J., Syder A.J., Yu Q.-C., Letai A., Paller A.S., Fuchs E.;
RT   "The genetic basis of epidermolytic hyperkeratosis: a disorder of
RT   differentiation-specific epidermal keratin genes.";
RL   Cell 70:811-819(1992).
RN   [7]
RP   VARIANTS EHK HIS-154; CYS-156; HIS-156; ASP-160 AND GLN-442.
RX   MEDLINE=94136477; PubMed=7508181;
RA   Chipev C.C., Yang J.-M., Digiovanna J.J., Steinert P.M., Marekov L.,
RA   Compton J.G., Bale S.J.;
RT   "Preferential sites in keratin 10 that are mutated in epidermolytic
RT   hyperkeratosis.";
RL   Am. J. Hum. Genet. 54:179-190(1994).
RN   [8]
RP   VARIANTS EHK PRO-156 AND SER-156.
RX   MEDLINE=94117870; PubMed=7507152;
RA   McLean W.H.I., Eady R.A., Dopping-Hepenstal P.J., McMillan J.R.,
RA   Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,
RA   Morley S.M.;
RT   "Mutations in the rod 1A domain of keratins 1 and 10 in bullous
RT   congenital ichthyosiform erythroderma (BCIE).";
RL   J. Invest. Dermatol. 102:24-30(1994).
RN   [9]
RP   VARIANTS.
RX   MEDLINE=92141228; PubMed=1371013;
RA   Korge B.P., Gan S.-Q., McBridge O.W., Mischke D., Steinert P.M.;
RT   "Extensive size polymorphism of the human keratin 10 chain resides in
RT   the C-terminal V2 subdomain due to variable numbers and sizes of
RT   glycine loops.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:910-914(1992).
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC       KERATIN 10 IS GENERALLY ASSOCIATED WITH KERATIN 1.
CC   -!- TISSUE SPECIFICITY: SEEN IN ALL SUPRABASAL CELL LAYERS INCLUDING
CC       STRATUM CORNEUM.
CC   -!- POLYMORPHISM: A NUMBER OF ALLELES ARE KNOWN THAT MAINLY DIFFER IN
CC       THE GLY-RICH REGION (POSITIONS 490-560).
CC   -!- DISEASE: DEFECTS IN KRT10 AND KRT1 ARE THE CAUSE OF EPIDERMOLYTIC
CC       HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL
CC       ICHTHYOSIFORM ERYTHRODERMA (BIE)); A HEREDITARY SKIN DISORDER
CC       CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
CC       CORNEUM.
CC   -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC       MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II
CC       (NEUTRAL TO BASIC; 56-70 KDA) [K1 TO K8].
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY IN
CC       POSITIONS 513 TO 555.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X14487; CAA32649.1; -.
DR   EMBL; M19156; AAA59468.1; -.
DR   EMBL; M77663; AAA59199.1; -.
DR   EMBL; L20218; AAB59438.1; -.
DR   EMBL; L20219; AAB59439.1; -.
DR   PIR; S02158; KRHU0.
DR   AARHUS/GHENT-2DPAGE; 7405; IEF.
DR   MIM; 148080; -.
DR   MIM; 113800; -.
DR   INTERPRO; IPR001664; -.
DR   INTERPRO; IPR002957; -.
DR   PFAM; PF00038; filament; 1.
DR   PRINTS; PR01248; TYPE1KERATIN.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin;
KW   Disease mutation; Polymorphism.
FT   DOMAIN        1    145       HEAD.
FT   DOMAIN      146    456       ROD.
FT   DOMAIN      457    593       TAIL.
FT   DOMAIN      146    181       COIL 1A.
FT   DOMAIN      182    202       LINKER 1.
FT   DOMAIN      203    294       COIL 1B.
FT   DOMAIN      295    317       LINKER 12.
FT   DOMAIN      318    456       COIL 2.
FT   DOMAIN        6    144       GLY/PHE/SER-RICH.
FT   DOMAIN      451    590       GLY/SER-RICH.
FT   VARIANT     154    154       N -> H (IN EHK).
FT                                /FTId=VAR_003826.
FT   VARIANT     156    156       R -> H (IN EHK).
FT                                /FTId=VAR_003827.
FT   VARIANT     156    156       R -> C (IN EHK).
FT                                /FTId=VAR_003828.
FT   VARIANT     156    156       R -> P (IN EHK).
FT                                /FTId=VAR_003829.
FT   VARIANT     156    156       R -> S (IN EHK).
FT                                /FTId=VAR_003830.
FT   VARIANT     160    160       Y -> D (IN EHK).
FT                                /FTId=VAR_003831.
FT   VARIANT     161    161       L -> S (IN EHK).
FT                                /FTId=VAR_003832.
FT   VARIANT     442    442       L -> Q (IN EHK).
FT                                /FTId=VAR_003833.
FT   CONFLICT    197    197       S -> G (IN REF. 3).
FT   CONFLICT    279    280       EL -> YV (IN REF. 2).
FT   CONFLICT    312    312       V -> I (IN REF. 2).
FT   CONFLICT    340    340       F -> V (IN REF. 2).
FT   CONFLICT    374    374       A -> R (IN REF. 2).
FT   CONFLICT    408    408       Q -> H (IN REF. 1).
FT   CONFLICT    451    451       S -> G (IN REF. 3).
FT   CONFLICT    460    461       GG -> RS (IN REF. 2).
FT   CONFLICT    477    477       S -> T (IN REF. 2).
FT   CONFLICT    482    482       S -> T (IN REF. 2).
FT   CONFLICT    487    490       MISSING (IN REF. 3).
FT   CONFLICT    503    503       S -> T (IN REF. 2).
FT   CONFLICT    508    508       S -> T (IN REF. 2).
FT   CONFLICT    519    527       MISSING (IN REF. 3).
FT   CONFLICT    513    518       YGGGSS -> LRGELH (IN REF. 2).
FT   CONFLICT    519    529       MISSING (IN REF. 2).
FT   CONFLICT    532    536       GGSSS -> AHST (IN REF. 2).
FT   CONFLICT    543    543       S -> N (IN REF. 2).
FT   CONFLICT    551    555       YGGGS -> LRGRH (IN REF. 2).
SQ   SEQUENCE   593 AA;  59519 MW;  36095FA0BAF985B5 CRC64;


  Query Match             78.7%;  Score 59;  DB 1;  Length 593;
  Best Local Similarity   78.6%;  Pred. No. 5.2;
  Matches   11;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           |||| || :|||||
```

```
Db    547 GGGGYGGGSSGGGS 560


RESULT  10
GRP2_NICSY
ID   GRP2_NICSY     STANDARD;      PRT;   214 AA.
AC   P27484;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 2 PRECURSOR.
GN   GRP-2.
OS   Nicotiana sylvestris (Wood tobacco).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC   Solanales; Solanaceae; Nicotiana.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92003709; PubMed=1912512;
RA   Obokata J., Ohme M., Hayashida N.;
RT   "Nucleotide sequence of a cDNA clone encoding a putative glycine-rich
RT   protein of 19.7 kDa in Nicotiana sylvestris.";
RL   Plant Mol. Biol. 17:953-955(1991).
CC   -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60007; CAA42622.1; -.
DR   PIR; S17731; KNNT2S.
DR   HSSP; P15277; 1MJC.
DR   INTERPRO; IPR001878; -.
DR   INTERPRO; IPR002059; -.
DR   PFAM; PF00313; CSD; 1.
DR   PFAM; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00050; COLDSHOCK.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PROSITE; PS00352; COLD_SHOCK; UNKNOWN_1.
KW   Cell wall; Structural protein; Repeat; Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   CHAIN         ?    214       GLYCINE-RICH CELL WALL STRUCTURAL
FT                                PROTEIN 2.
FT   DOMAIN       82    158       GLY-RICH.
FT   DOMAIN      176    195       GLY-RICH.
SQ   SEQUENCE   214 AA;  19746 MW;  E28DB84538F2A0AA CRC64;


  Query Match             77.3%;  Score 58;  DB 1;  Length 214;
  Best Local Similarity   78.6%;  Pred. No. 2.8;
  Matches   11;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           |||| ||  |||||
Db     144 GGGGYGGGGSGGGS 157


RESULT  11
SIX3_MOUSE
ID   SIX3_MOUSE     STANDARD;      PRT;   333 AA.
AC   Q62233; P70176; P70177;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HOMEOBOX PROTEIN SIX3.
GN   SIX3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=EMBRYONIC BRAIN;
RX   MEDLINE=96125147; PubMed=8575305;
RA   Oliver G., Mailhos A., Wehr R., Copeland N.G., Jenkins N.A.,
RA   Gruss P.;
RT   "Six3, a murine homologue of the sine oculis gene, demarcates the
RT   most anterior border of the developing neural plate and is expressed
RT   during eye development.";
RL   Development 121:4045-4055(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=96409319; PubMed=8814301;
RA   Kawakami K., Ohto H., Takizawa T., Saito T.;
RT   "Identification and expression of Six family genes in mouse retina.";
RL   FEBS Lett. 393:259-263(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN VISUAL SYSTEM DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SIX3A AND SIX3B (SHOWN HERE);
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT E6.5 OF EMBRYO DEVELOPMENT
CC       AROUND THE ANTERIOR BORDER. AT E8.5, EXPRESSION IS FOUND OVER THE
CC       ANTERIOR NEURAL PLATE. AT E9.5, IN THE DIENCEPHALIC PART OF THE
CC       VENTRAL FOREBRAIN, OPTIC VESICLES, OLFACTORY PLACODES AND RATHKE'S
CC       POUCH. IN LATER STAGES, PRESENT IN HYPOTHALAMUS, EYES AND
CC       PITUITARY.
CC   -!- SIMILARITY: BELONGS TO THE SIX/SINE OCULIS FAMILY OF HOMEODOMAIN
CC       PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X90871; CAA62379.1; ALT_INIT.
DR   EMBL; D83144; BAA11822.1; -.
DR   EMBL; D83145; BAA11823.1; -.
DR   MGD; MGI:102764; SIX3.
DR   INTERPRO; IPR001356; -.
DR   PFAM; PF00046; homeobox; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; FALSE_NEG.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Developmental protein; Homeobox; DNA-binding; Nuclear protein;
KW   Alternative splicing.
FT   DOMAIN       40     51       POLY-GLY.
FT   DOMAIN       55     65       POLY-GLY.
FT   DOMAIN       60     70       POLY-GLY.
FT   DNA_BIND    207    266       HOMEOBOX.
FT   DOMAIN      264    267       POLY-ALA.
FT   VARSPLIC    271    286       LQHQAIGPSGMRSLAE -> SVAGTAARPPQA (IN
FT                                ISOFORM SIX3A).
FT   VARSPLIC    287    333       MISSING (IN ISOFORM SIX3A).
FT   CONFLICT     44     44       G -> GG (IN REF. 1).
FT   CONFLICT    118    119       VA -> WP (IN REF. 1).
FT   CONFLICT    278    333       PSGMRSLAEPGCPTHGSAESPSTAASPTTSVSSLTERADTG
FT                                TSILSVTSSDSECDV -> ERDALPGRARLPHARLSRVTVH
FT                                GGQPDHQCVQPDGARGHRHFDPLGNLQRLGM (IN REF.
FT                                1).
SQ   SEQUENCE   333 AA;  35592 MW;  1AD7D3C4388043B9 CRC64;


  Query Match             77.3%;  Score 58;  DB 1;  Length 333;
  Best Local Similarity   71.4%;  Pred. No. 4.1;
  Matches   10;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           |||| | || |||:
Db      46 GGGGGGNRAGGGGA 59
```

```
RESULT  12
DSXF_DROME
ID   DSXF_DROME     STANDARD;      PRT;   427 AA.
AC   P23022;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DOUBLESEX PROTEIN, FEMALE-SPECIFIC.
GN   DSX.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PUPAE, AND LARVA;
RX   MEDLINE=89168451; PubMed=2493994;
RA   Burtis K.C., Baker B.S.;
RT   "Drosophila doublesex gene controls somatic sexual differentiation by
RT   producing alternatively spliced mRNAs encoding related sex-specific
RT   polypeptides.";
RL   Cell 56:997-1010(1989).
RN   [2]
RP   DNA-BINDING.
RX   MEDLINE=91330881; PubMed=1907913;
RA   Burtis K.C., Coschigano K.T., Baker B.S., Wensink P.C.;
RT   "The doublesex proteins of Drosophila melanogaster bind directly to a
RT   sex-specific yolk protein gene enhancer.";
RL   EMBO J. 10:2577-2582(1991).
RN   [3]
RP   DNA-BINDING DOMAIN, AND MUTAGENESIS.
RX   MEDLINE=93178426; PubMed=8440242;
RA   Erdman S.E., Burtis K.C.;
RT   "The Drosophila doublesex proteins share a novel zinc finger related
RT   DNA binding domain.";
RL   EMBO J. 12:527-535(1993).
CC   -!- FUNCTION: CONTROLS SOMATIC SEXUAL DIFFERENTIATION. BINDS DIRECTLY
CC       AND SPECIFICALLY TO THE ENHANCER FBE (FAT BODY ENHANCER) OF
CC       THE YOLK PROTEIN 1 AND 2 GENES (YP1 AND YP2). THIS ENHANCER IS
CC       SUFFICIENT TO DIRECT THE FEMALE-SPECIFIC TRANSCRIPTION
CC       CHARACTERISTIC OF THE YP GENES IN ADULT FAT BODIES.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A FEMALE-SPECIFIC ISOFORM (SHOWN
CC       HERE) AND A MALE-SPECIFIC ISOFORM (AC P23023); ARE PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- MISCELLANEOUS: EXPERIMENTALLY SHOWN TO BIND ZINC.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25292; AAA17840.1; -.
DR   EMBL; M25293; AAA17841.1; -.
DR   PIR; A32372; A32372.
DR   HSSP; P04002; 1ATF.
DR   TRANSFAC; T00955; -.
DR   FLYBASE; FBgn0000504; dsx.
DR   INTERPRO; IPR001275; -.
DR   PFAM; PF00751; DM-domain; 1.
KW   Sexual differentiation; Alternative splicing; DNA-binding;
KW   Transcription regulation; Nuclear protein; Zinc-finger.
FT   DOMAIN        1    397       COMMON TO MALE AND FEMALE FORM.
FT   DNA_BIND     39    104
FT   DOMAIN      119    224       HIS-RICH.
FT   DOMAIN      267    296       SER/GLY-RICH.
FT   MUTAGEN      47     47       C->A,H: ABOLISHES DNA-BINDING.
FT   MUTAGEN      50     50       H->Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      59     59       H->Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      68     68       C->D,Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      70     70       C->Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      91     91       R->Q: ABOLISHES DNA-BINDING.
SQ   SEQUENCE   427 AA;  44768 MW;  75143686AC717315 CRC64;


  Query Match             77.3%;  Score 58;  DB 1;  Length 427;
  Best Local Similarity   71.4%;  Pred. No. 5;
  Matches   10;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           |||| || :||||:
Db     274 GGGGGGGGSSGGGA 287


RESULT  13
LORI_MOUSE
ID   LORI_MOUSE     STANDARD;      PRT;   481 AA.
AC   P18165;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   LORICRIN.
GN   LOR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90275605; PubMed=2190691;
RA   Mehrel T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
RA   Cheng C., Lichti U., Bisher M.E., Steven A.C., Steinart P.M.,
RA   Yuspa S.H., Roop D.R.;
RT   "Identification of a major keratinocyte cell envelope protein,
RT   loricrin.";
RL   Cell 61:1103-1112(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=95256248; PubMed=7738016;
RA   DiSepio D., Jones A., Longley M.A., Bundman D., Rothnagel J.A.,
RA   Roop D.R.;
RT   "The proximal promoter of the mouse loricrin gene contains a
RT   functional AP-1 element and directs keratinocyte-specific but not
RT   differentiation-specific expression.";
RL   J. Biol. Chem. 270:10792-10799(1995).
CC   -!- FUNCTION: MAJOR KERATINOCYTE CELL ENVELOPE PROTEIN.
CC   -!- SUBUNIT: MONOMERS ARE CROSSLINKED BY DISULFIDE AND N-(GAMMA-
CC       GLUTAMYL) LYSINE ISODIPEPTIDE BONDS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34398; AAA39444.1; -.
DR   EMBL; U09189; AAA82152.1; -.
DR   PIR; A35628; A35628.
DR   HSSP; P10968; 1WGC.
DR   MGD; MGI:96816; LOR.
KW   Keratinocyte.
SQ   SEQUENCE   481 AA;  37830 MW;  97349A786FF239FE CRC64;


  Query Match             77.3%;  Score 58;  DB 1;  Length 481;
  Best Local Similarity   78.6%;  Pred. No. 5.6;
  Matches   11;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           |||| ||  |||||
```

```
Db      22 GGGGGGGYYSGGGS 35


RESULT  14
DSXM_DROME
ID   DSXM_DROME     STANDARD;      PRT;   549 AA.
AC   P23023; Q9VHY0;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   DOUBLESEX PROTEIN, MALE-SPECIFIC.
GN   DSX.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PUPAE, AND LARVA;
RX   MEDLINE=89168451; PubMed=2493994;
RA   Burtis K.C., Baker B.S.;
RT   "Drosophila doublesex gene controls somatic sexual differentiation by
RT   producing alternatively spliced mRNAs encoding related sex-specific
RT   polypeptides.";
RL   Cell 56:997-1010(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   DNA-BINDING.
RX   MEDLINE=91330881; PubMed=1907913;
RA   Burtis K.C., Coschigano K.T., Baker B.S., Wensink P.C.;
RT   "The doublesex proteins of Drosophila melanogaster bind directly to a
RT   sex-specific yolk protein gene enhancer.";
RL   EMBO J. 10:2577-2582(1991).
RN   [4]
RP   DNA-BINDING DOMAIN, AND MUTAGENESIS.
RX   MEDLINE=93178426; PubMed=8440242;
RA   Erdman S.E., Burtis K.C.;
RT   "The Drosophila doublesex proteins share a novel zinc finger related
RT   DNA binding domain.";
RL   EMBO J. 12:527-535(1993).
CC   -!- FUNCTION: CONTROLS SOMATIC SEXUAL DIFFERENTIATION. BINDS DIRECTLY
CC       AND SPECIFICALLY TO THE ENHANCER FBE (FAT BODY ENHANCER) OF
CC       THE YOLK PROTEIN 1 AND 2 GENES (YP1 AND YP2). THIS ENHANCER IS
CC       SUFFICIENT TO DIRECT THE FEMALE-SPECIFIC TRANSCRIPTION
CC       CHARACTERISTIC OF THE YP GENES IN ADULT FAT BODIES.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A MALE-SPECIFIC ISOFORM (SHOWN
CC       HERE) AND A FEMALE-SPECIFIC ISOFORM (AC P23022); ARE PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- MISCELLANEOUS: EXPERIMENTALLY SHOWN TO BIND ZINC.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M25292; AAA17840.1; -.
DR   EMBL; M25294; AAA17842.1; -.
DR   EMBL; AE003676; AAF54168.1; -.
DR   PIR; B32372; B32372.
DR   HSSP; P04002; 1ATF.
DR   TRANSFAC; T00956; -.
DR   FLYBASE; FBgn0000504; dsx.
DR   INTERPRO; IPR001275; -.
DR   PFAM; PF00751; DM-domain; 1.
KW   Sexual differentiation; Alternative splicing; DNA-binding;
KW   Transcription regulation; Nuclear protein; Zinc-finger.
FT   DOMAIN        1    397       COMMON TO MALE AND FEMALE FORM.
FT   DNA_BIND     39    104
FT   DOMAIN      119    224       HIS-RICH.
FT   DOMAIN      267    296       SER/GLY-RICH.
FT   MUTAGEN      47     47       C->A,H: ABOLISHES DNA-BINDING.
FT   MUTAGEN      50     50       H->Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      59     59       H->Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      68     68       C->D,Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      70     70       C->Y: ABOLISHES DNA-BINDING.
FT   MUTAGEN      91     91       R->Q: ABOLISHES DNA-BINDING.
SQ   SEQUENCE   549 AA;  57409 MW;  3C1B92724E4CE083 CRC64;


  Query Match             77.3%;  Score 58;  DB 1;  Length 549;
  Best Local Similarity   71.4%;  Pred. No. 6.2;
  Matches   10;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           |||| || :||||:
Db     274 GGGGGGGGSSGGGA 287


RESULT  15
RLA3_ORYSA
ID   RLA3_ORYSA     STANDARD;      PRT;   118 AA.
AC   P56724;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   60S ACIDIC RIBOSOMAL PROTEIN P3 (P1/P2-LIKE).
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95078950; PubMed=7987417;
```

```
RA   Sasaki T., Song J., Koga-Ban Y., Matsui E., Fang F., Higo H.,
RA   Nagasaki H., Hori M., Miya M., Murayama-Kayano E.;
RT   "Toward cataloguing all rice genes: large-scale sequencing of
RT   randomly chosen rice cDNAs from a callus cDNA library.";
RL   Plant J. 6:615-624(1994).
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC       PROTEIN SYNTHESIS.
CC   -!- PTM: PHOSPHORYLATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D15754; -; NOT_ANNOTATED_CDS.
KW   Ribosomal protein; Phosphorylation.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   118 AA;  11646 MW;  6FB60E57532C1CB4 CRC64;


  Query Match             76.0%;  Score 57;  DB 1;  Length 118;
  Best Local Similarity   71.4%;  Pred. No. 2.1;
  Matches   10;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGGSGGRASGGGS 14
           |||| || |:|||:
Db      69 GGGGGGGFAAGGGA 82


Search completed: March 15, 2001, 11:11:12
Job time: 1094 sec
```